

Interference Mitigation Using Asynchronous Transmission and Sampling Diversity

Mehdi Ganji and Hamid Jafarkhani

Abstract—In this paper, we show that by investigating inherent time delays between different users in a multiuser scenario, we are able to cancel interference more efficiently. Time asynchrony provides another tool to cancel interference which results in preserving other resources like frequency, time and code. Therefore, we can save the invaluable resource of frequency band and also increase spectral efficiency. A sampling method is presented which results in independent noise samples and obviates the need for the complex process of noise whitening. By taking advantage of this sampling method and its unique structure, we implement maximum-likelihood sequence detection which outperforms synchronous maximum-likelihood detection. We also present successive interference cancellation with hard decision passing which gives rise to a novel forward-backward belief propagation method. Next, the performance of zero forcing detection is analyzed. Simulation results are also presented to verify our analysis.

I. INTRODUCTION

There are many applications where multiple users share a common channel to transmit data to a receiver. Numerous examples of multiaccess communication include uplink transmission of a single cell in a cellular system, a group of twisted-pair copper subscriber lines transmitting data to the same switching office, multiple ground stations communicating with a satellite and interactive cable television networks. The key challenge in multiuser transmissions or multiple access channels is Interuser Interference. Over several decades, many methods have been introduced to address this problem [1], [2]. Most of these methods are based on assigning orthogonal dimensions to different users to be able to separate them and prevent interference. For example, time division multiple access (TDMA) protocols allocate different time slots to different users to mitigate interference. The same concept can be applied by partitioning the frequency spectrum among different users, which is called frequency division multiple access (FDMA). Code division multiple access is another scheme used to surpass interuser interference in which users are multiplexed by distinct codes rather than by orthogonal frequency bands, or by orthogonal time slots [3]. More recently, multiple receive antennas are utilized at the receive side to take advantage of the spatial domain in order to cancel interference [4], [5].

In this paper, we investigate the timing mismatch between users as an additional resource to address the problem of interuser interference. By exploiting time delays between users

and employing an appropriate sampling method, we design detection methods which not only cancel the interference effectively, but also outperform the synchronous ones. When timing mismatch is used to cancel the interuser interference, resources like frequency spectrum, time and receive antenna can be employed to improve the performance. There are other examples in the literature in which asynchronous transmission outperforms synchronous transmission. For example, by using timing delays between users, zero forcing (ZF) detection can be performed with one receive antenna and additional receive antennas can be used to gain diversity [6], [7]. In [6], the authors proposed a ZF receiver in MIMO setting which takes advantage of timing mismatch between data streams and provides full diversity of M , where M is the number of receive antennas. However, a crucial impairment of their receiver design is addressed in [7]. The design of asynchronous differential decoding methods which outperform their synchronous counterparts is discussed in [8], [9]. In this paper, we present sampling diversity and provide several decoders to gain advantages from asynchronous transmission. We analytically prove that our ZF method provides full diversity and we study its asymptotic performance for large number of receive antennas.

II. SYSTEM MODEL

A. General Settings

We consider a system with K users, transmitting data to a common receiver simultaneously, which can have one receive antenna or multiple ones. Due to different physical locations of users, their signal is received with various time delays. It is assumed that each data stream is received with an arbitrary delay smaller than the symbol interval and only the receiver knows the time delays. The signal transmitted from User k is described by:

$$s_k(t) = \sum_{i=1}^N b_k(i)p(t - (i-1)T_s) \quad (1)$$

where T_s is the symbol length and $p(\cdot)$ is the pulse-shaping filter with non-zero duration of T . Also, N is the frame length and $b_k(i)$ is the transmitted symbol by User k in the i th time slot. The transmitted signals are received with a relative delay of τ_k and a channel path gain of h_k . Then, the received signal can be represented by:

$$y(t) = \sum_{k=1}^K h_k s_k(t - \tau_k) + n(t) \quad (2)$$

where K is the number of users and $n(t)$ is the white noise with variance of σ^2 . Without loss of generality, we assume that $0 = \tau_1 < \tau_2 < \dots < \tau_K < T$.

This work was supported in part by the NSF Award CCF-1526780. The authors are with the Center for Pervasive Communications and Computing, University of California, Irvine, CA 92697-2625 USA (e-mail: {mganji, hamidj}@uci.edu).

B. Output Samples

The output of the matched filter at each receiver antenna can be sampled at different sampling times associated with different users as shown in Fig. 1a. These sets of samples provide sufficient statistics for decoding transmitted symbols [10]. We can break down the integrals corresponding to the sampling in Fig. 1a to define a new sampling method as shown in Fig. 1b. The corresponding output samples are written in Eq. (9) where τ_{k+1} is an auxiliary variable equal to T .

By defining intermediate variables $u_{ji}(l, k)$ and noise samples

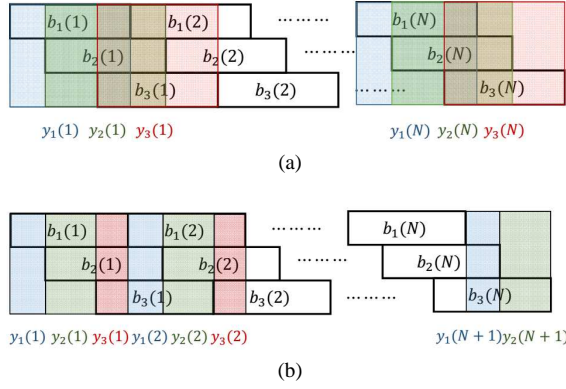


Fig. 1: Sampling methods

$v_l(j)$ in Eqs. (10) and (11), at the bottom of the page, we can write output samples in a more compact way:

$$y_l(j) = \sum_{k=1}^K \sum_{i=1}^N b_k(i) u_{ji}(l, k) h_k + v_l(j) \quad (3)$$

$j = 1, \dots, N+1$ sampling time index
 $l = 1, \dots, K$ index of matched user

Defining $\mathbf{y}(j) = [y_1(j), y_2(j), \dots, y_K(j)]^T$ and $\mathbf{b}(i) = [b_1(i), b_2(i), \dots, b_K(i)]^T$, then, $\mathbf{y}(j)$ for different values of j can be written as:

$$\mathbf{y}(j) = \sum_{i=1}^N \mathbf{U}_{ji} \mathbf{h} \mathbf{b}(i) + \mathbf{v}(j) \quad 1 \leq j \leq N+1 \quad (4)$$

where $\mathbf{h} = \text{diag}[h_1, h_2, \dots, h_K]$, $\mathbf{v}(j) = [v_1(j), v_2(j), \dots, v_K(j)]^T$ and \mathbf{U}_{ji} is a $K \times K$ matrix

whose elements are defined as $U_{ji}(l, k) = u_{ji}(l, k)$. The next step is to put all vectors of $\mathbf{y}(j)$ together and define \mathbf{y} as $[\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(N+1)]^T$. Then, \mathbf{y} can be written as:

$$\mathbf{y} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} & \mathbf{U}_{13} & \dots & \mathbf{U}_{1N} \\ \mathbf{U}_{21} & \mathbf{U}_{11} & \mathbf{U}_{12} & \dots & \mathbf{U}_{1(N-1)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{U}_{(N-1)1} & \dots & \mathbf{U}_{21} & \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{N1} & \dots & \mathbf{U}_{31} & \mathbf{U}_{21} & \mathbf{U}_{11} \\ \mathbf{U}_{(N+1)1} & \dots & \mathbf{U}_{41} & \mathbf{U}_{31} & \mathbf{U}_{21} \end{pmatrix} \begin{pmatrix} \mathbf{h} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{h} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{h} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{h} \end{pmatrix} \begin{pmatrix} \mathbf{b}(1) \\ \mathbf{b}(2) \\ \vdots \\ \mathbf{b}(N) \end{pmatrix} + \mathbf{v}$$

$$= \mathbf{U} \mathbf{H} \mathbf{b} + \mathbf{v} \quad (5)$$

Block Toeplitz structure of \mathbf{U} originates from the fact that $u_{(j+m)(i+m)}(l, k) = u_{ji}(l, k)$. This can be verified by a change of variable in Eq. (10). Based on the relation between T and T_s , different numbers of adjacent symbols interfere with each other. For example, for rectangular pulse shapes, i.e., $T = T_s$, at each instant only current and previous symbols cause interference. In other words, only \mathbf{U}_{11} and \mathbf{U}_{21} are nonzero. Without loss of generality, we assume that $T = 1$, therefore \mathbf{U}_{11} and \mathbf{U}_{21} are defined as follows:

$$\mathbf{U}_{11} = \begin{pmatrix} \tau_2 - \tau_1 & 0 & \dots & 0 \\ \tau_3 - \tau_2 & \tau_3 - \tau_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_K - \tau_{K-1} & \dots & \tau_K - \tau_{K-1} & 0 \\ 1 - \tau_K & \dots & 1 - \tau_K & 1 - \tau_K \end{pmatrix} \quad (6)$$

$$\mathbf{U}_{21} = \begin{pmatrix} 0 & \tau_2 - \tau_1 & \dots & \tau_2 - \tau_1 \\ 0 & 0 & \dots & \tau_3 - \tau_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \tau_{K-1} - \tau_K \\ 0 & \dots & 0 & 0 \end{pmatrix} \quad (7)$$

Hence, for rectangular pulse shapes, the system model simplifies to:

$$\mathbf{y} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{U}_{21} & \mathbf{U}_{11} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{U}_{21} & \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{U}_{21} & \mathbf{U}_{11} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{U}_{21} \end{pmatrix} \begin{pmatrix} \mathbf{h} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{h} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{h} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{h} \end{pmatrix} \begin{pmatrix} \mathbf{b}(1) \\ \mathbf{b}(2) \\ \vdots \\ \mathbf{b}(N) \end{pmatrix} + \mathbf{v}$$

$$= \mathbf{U} \mathbf{H} \mathbf{b} + \mathbf{v} \quad (8)$$

The important fact about this sampling method is that the covariance matrix of noise samples is diagonal. With a small abuse of notation, we denote $\text{Diag}(\mathbf{U}_{11})$ as a diagonal

$$y_l(j) = \int_{\tau_l + (j-1)T_s}^{\tau_{l+1} + (j-1)T_s} \sum_{k=1}^K \sum_{i=1}^N b_k(i) p(t - (j-1)T_s - \tau_l) p(t - (i-1)T_s - \tau_k) h_k dt$$

$$+ \int_{\tau_l + (j-1)T_s}^{\tau_{l+1} + (j-1)T_s} n(t) p(t - (j-1)T_s - \tau_l) dt \quad 1 \leq l \leq K, 1 \leq j \leq N+1 \quad (9)$$

$$u_{ji}(l, k) = \int_{\tau_l + (j-1)T_s}^{\tau_{l+1} + (j-1)T_s} p(t - (j-1)T_s - \tau_l) p(t - (i-1)T_s - \tau_k) dt \quad (10)$$

$$v_l(j) = \int_{\tau_l + (j-1)T_s}^{\tau_{l+1} + (j-1)T_s} n(t) p(t - (j-1)T_s - \tau_l) dt \quad (11)$$

matrix including diagonal elements of \mathbf{U}_{11} . Then, it can be shown that $E[\mathbf{v}\mathbf{v}^H]$ is equal to $\sigma^2(\mathbf{I}_N \otimes \text{Diag}(\mathbf{U}_{11}))$, where \mathbf{I}_N is an $N \times N$ identity matrix and (\otimes) is Kronecker product.

Since the statistically sufficient samples in Fig. 1a can be created from samples in Fig. 1b, the samples in Fig. 1b, i.e. Eq. (8), are sufficient statistics too. Both of these sampling methods introduce intentional intersymbol interference (ISI) and impose memory on the system; however, they have some differences:

- 1) Sampling intervals in Fig. 1b are smaller and need faster sampler.
- 2) Since sampling intervals are disjoint in Fig. 1b, noise samples are independent. However, due to sampling overlap, the noise samples in Fig. 1a are correlated.
- 3) The sampling in Fig. 1b results in an overdetermined system, while the number of output samples in Fig. 1a is equal to the number of input symbols.

For rectangular pulse shape, the input-output relationship of the sampling method in Fig. 1a is:

$$\mathbf{y} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{R}_{21} & \mathbf{R}_{11} & \mathbf{R}_{12} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{R}_{21} & \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{R}_{21} & \mathbf{R}_{11} \end{pmatrix} \begin{pmatrix} h & 0 & 0 & \dots & 0 \\ 0 & h & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h & 0 \\ 0 & \dots & 0 & 0 & h \end{pmatrix} \begin{pmatrix} b(1) \\ b(2) \\ \vdots \\ b(N) \end{pmatrix} + \mathbf{n} \quad (12)$$

$$= \mathbf{R}\mathbf{H}\mathbf{b} + \mathbf{n}$$

where \mathbf{R}_{11} , \mathbf{R}_{21} and \mathbf{R}_{12} are defined as:

$$\mathbf{R}_{11} = \begin{pmatrix} 1 & 1-(\tau_2-\tau_1) & \dots & 1-(\tau_K-\tau_1) \\ 1-(\tau_2-\tau_1) & 1 & \dots & 1-(\tau_K-\tau_2) \\ \vdots & \ddots & \ddots & \vdots \\ 1-(\tau_{K-1}-\tau_1) & \dots & 1 & 1-(\tau_K-\tau_{K-1}) \\ 1-(\tau_K-\tau_1) & \dots & 1-(\tau_K-\tau_{K-1}) & 1 \end{pmatrix} \quad (13)$$

$$\mathbf{R}_{12} = (\mathbf{R}_{21})^T = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \tau_2-\tau_1 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ \tau_{(K-1)}-\tau_1 & \tau_{(K-1)}-\tau_2 & \dots & 0 & 0 \\ \tau_K-\tau_1 & \tau_K-\tau_2 & \dots & \tau_K-\tau_{K-1} & 0 \end{pmatrix} \quad (14)$$

Because of intersection between sampling intervals, noise samples are correlated and noise whitening procedure needs to be performed before symbol detection. Noise whitening involves Cholesky decomposition and matrix inversion which increases complexity of receiver.

III. RECEIVER DESIGN

In this section we introduce different detection methods which take advantage of distinct features of the sampling method shown in Fig. 1b. One of these features is converting a memoryless system into a system with memory and independent noise samples. This enables us to implement the Viterbi algorithm based on samples in Eq. (8). The other feature is that this sampling method provides extra output samples

which can be used to improve detection methods. For example, these extra samples make it possible to carry out successive interference cancellation (SIC) backward and forward. Also, by means of introduced ISI, zero forcing detection can be performed even with one receive antenna, which is impossible in synchronous multiuser transmission. In what follows, we will show how asynchronous multiuser transmission can outperform synchronous multiuser transmission.

A. Maximum-Likelihood Sequence Detection (MLSD)

Due to inherent memory in the system that results from time delays, we can use the maximum-likelihood sequence detection method implemented by the Viterbi algorithm. The objective of maximum-likelihood sequence detector is to find the input sequence that maximizes the conditional probability, or the likelihood of the given output sequence. Exhaustive search over 2^{NK} different input sequences is an obvious choice, but it is impractical even for a moderate number of K and N . Fortunately, using the Viterbi algorithm, MLSD can be implemented by complexity order of 2^K [11]. For using the Viterbi algorithm, the likelihood metric should be additive and the noise samples should be independent. Therefore, the sampling method in Fig. 1b is the best fit for implementing the Viterbi algorithm. It reduces the complexity by avoiding the noise whitening procedure which involves Cholesky Decomposition and matrix inversion. We show that by using this sampling method, we can outperform synchronous ML detection with the same complexity order of 2^K .

Based on the recursive relation between input and output which is described as:

$$\mathbf{y}(j) = \mathbf{U}_{11}\mathbf{h}\mathbf{b}(j) - \mathbf{U}_{21}\mathbf{h}\mathbf{b}(j-1) + \mathbf{v}(j) \quad 2 \leq j \leq N$$

the trellis diagram of the system includes A^K states with A^K outgoing paths to the next states, and A^K incoming paths from previous states, where A is the size of the transmitted modulation. To calculate the metric for each path, we need to calculate the likelihood function as follows:

$$\begin{aligned} Pr(\mathbf{y}(j)|\mathbf{b}(j), \mathbf{b}(j-1)) &= \\ Pr(\mathbf{v}(j) = \mathbf{y}(j) - \mathbf{U}_{11}\mathbf{h}\mathbf{b}(j) - \mathbf{U}_{21}\mathbf{h}\mathbf{b}(j-1)) &= \\ \frac{1}{\sqrt{(2\pi)^K |\Sigma|}} \exp\left(-\frac{1}{2} \mathbf{s}_j^H \Sigma^{-1} \mathbf{s}_j\right) \end{aligned}$$

where $\mathbf{s}_j = \mathbf{y}(j) - \mathbf{U}_{11}\mathbf{h}\mathbf{b}(j) - \mathbf{U}_{21}\mathbf{h}\mathbf{b}(j-1)$ and $\Sigma = E[\mathbf{v}(j)\mathbf{v}(j)^H]$. By discarding common terms and simple calculations, the metric for each path can be defined as $\sum_{i=1}^K \frac{|\mathbf{s}_j(i)|^2}{\sigma^2 \mathbf{U}_{11}(i,i)}$. After calculating the path metrics, the final goal is to find the surviving path and trace it back to decode the transmitted symbols. The simulation result for this algorithm and its comparison with the synchronous ML detection is presented in Section V.

B. Successive Interference Cancellation with Hard Decision Passing

Despite the excellent performance provided by MLSD, its complexity grows exponentially with the number of users,

which might be prohibitive in some practical scenarios. Successive interference cancellation (SIC) detection that takes a serial approach to cancel interference can be used to reduce complexity. Using the sampling method in Fig. 1b, this serial approach can be either a forward SIC initiated from the first transmitted symbol, i.e., $b_1(1)$, or a backward SIC started from the last transmitted symbol, i.e., $b_K(N)$.

For example, for forward processing, $b_1(1)$ can be decoded by using $y_1(1)$ without interference, then $b_2(1)$ can be decoded by cancelling the interference of $b_1(1)$ from $y_2(1)$, and so on. The same procedure can be performed backwards. One can also combine forward and backward operations. However, when hard decisions are used, such a combination will not result in a noticeable gain. On the other hand, by using soft decisions, combining the forward and backward operations will improve the results as explained in the following section.

C. Forward Backward Belief Propagation Detection

In the previous section we introduced an SIC method which was performed by passing hard decisions of previously decoded symbols to cancel the interference. In this section, we introduce a similar detection method which passes likelihood values. By using likelihood values, instead of hard decisions, performance can be improved as shown by simulation results. Additionally, this method provides the opportunity to exploit benefits of backward processing as well. We explain the strategy of decoding for BPSK modulation and $K = 2$, but it can be also generalized to other modulations and other values of K . We also assume that transmitted symbols have the same prior probabilities and calculate the conditional probabilities as follows:

$$\begin{cases} a = P(y_1(1)|b_1(1) = 0) = \frac{1}{\sqrt{2\pi\rho_1\sigma^2}} \exp\left(-\frac{|y_1(1)+h_1\rho_1|^2}{2\rho_1\sigma^2}\right) \\ b = P(y_1(1)|b_1(1) = 1) = \frac{1}{\sqrt{2\pi\rho_1\sigma^2}} \exp\left(-\frac{|y_1(1)-h_1\rho_1|^2}{2\rho_1\sigma^2}\right) \\ P_0^{fw}(b_1(1)) = P(b_1(1) = 0|y_1(1)) = \frac{a}{a+b} \\ P_1^{fw}(b_1(1)) = P(b_1(1) = 1|y_1(1)) = \frac{b}{a+b} \\ c = P(y_2(1)|b_2(1) = 0, y_1(1)) = \\ P_0^{fw}(b_1(1)) \frac{1}{\sqrt{2\pi\rho_2\sigma^2}} \exp\left(-\frac{|y_2(1)+h_1\rho_1+h_2\rho_2|^2}{2\rho_2\sigma^2}\right) + \\ P_1^{fw}(b_1(1)) \frac{1}{\sqrt{2\pi\rho_2\sigma^2}} \exp\left(-\frac{|y_2(1)-h_1\rho_1+h_2\rho_2|^2}{2\rho_2\sigma^2}\right) \\ d = P(y_2(1)|b_2(1) = 1, y_1(1)) = \\ P_0^{fw}(b_1(1)) \frac{1}{\sqrt{2\pi\rho_2\sigma^2}} \exp\left(-\frac{|y_2(1)+h_1\rho_1-h_2\rho_2|^2}{2\rho_2\sigma^2}\right) + \\ P_1^{fw}(b_1(1)) \frac{1}{\sqrt{2\pi\rho_2\sigma^2}} \exp\left(-\frac{|y_2(1)-h_1\rho_1-h_2\rho_2|^2}{2\rho_2\sigma^2}\right) \\ P_0^{fw}(b_2(1)) = P(b_2(1) = 0|y_1(1), y_2(1)) = \frac{c}{c+d} \\ P_1^{fw}(b_2(1)) = P(b_2(1) = 1|y_1(1), y_2(1)) = \frac{d}{c+d} \end{cases}$$

where $\rho_i = U_{11}(i, i)$. Using these successive calculations, $P_0^{fw}(b_k(n))$ and $P_1^{fw}(b_k(n))$ can be found for all values of $1 \leq n \leq N$ and $1 \leq k \leq K$. As explained before, due to the structure of the sampling method in Fig. 1b, the last transmitted symbol can also be detected without interference and the same procedure can be applied backward to find $P_0^{bw}(b_k(n))$ and $P_1^{bw}(b_k(n))$. Using either of these likelihood

sets as a detection metric will result in an improvement over the hard-decision SIC method that was presented in the previous section. Moreover, the performance can even surpass the performance of the synchronous ML detection if we use forward and backward operations together and define the detection metric as:

$$\begin{aligned} P_0(b_k(n)) &= P_0^{fw}(b_k(n))P_0^{bw}(b_k(n)) \\ P_1(b_k(n)) &= P_1^{fw}(b_k(n))P_1^{bw}(b_k(n)) \end{aligned}$$

Simulation results are presented in Section V.

D. Zero Forcing

One of the well-known linear multiuser receivers is the ZF receiver which cancels the interference caused by the other users in the expense of enhancing the noise. In a synchronized system, we need at least K receive antennas to be able to perform ZF detection; however, by exploiting asynchrony, we can perform ZF with only one receive antenna [6], [7]. To have a fair comparison with the synchronous case, we study the system model when multiple receive antennas are used at the receiver. By stacking output samples of all receive antennas together we can represent the system model as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} U & 0 & \dots & 0 \\ 0 & U & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & U \end{bmatrix} \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_M \end{bmatrix} \mathbf{b} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_M \end{bmatrix}$$

$$\mathbf{y}_{tot} = \mathbf{U}_{tot} \mathbf{H}_{tot} \mathbf{b} + \mathbf{v}_{tot}$$

$$\mathbf{y}_{tot} = \mathbf{L}_{tot} \mathbf{b} + \mathbf{v}_{tot}$$

where M is number of receive antennas. Then, the zero-forcing detector is defined as:

$$\tilde{\mathbf{y}} = (\mathbf{L}_{tot}^H \mathbf{\Sigma}_{tot}^{-1} \mathbf{L}_{tot})^{-1} \mathbf{L}_{tot}^H \mathbf{\Sigma}_{tot}^{-1} \mathbf{y}_{tot} = \mathbf{b} + \tilde{\mathbf{v}} \quad (15)$$

where $\mathbf{\Sigma}_{tot} = \mathbf{I}_M \otimes \mathbf{\Sigma}$, and $\mathbf{\Sigma}$ is an $NK \times NK$ diagonal matrix representing the covariance matrix of noise samples \mathbf{v}_i , $1 \leq i \leq M$. The noise enhancement factor is $(\mathbf{L}_{tot}^H \mathbf{\Sigma}_{tot}^{-1} \mathbf{L}_{tot})^{-1}$, which affects the receiver performance and will be studied in the following section.

IV. PERFORMANCE ANALYSIS

All asynchronous receivers presented in the previous section provide full diversity. Because the ZF receiver has the worst performance among all introduced receivers, we only need to show full diversity for the ZF receiver. The system represented in Eq. (15) consists of NK subchannels, each of them having SNR of $\frac{E[|b_k(i)|^2]}{\mathbf{COV}_{\tilde{\mathbf{v}}}(i, i)}$, $1 \leq i \leq NK$, where $\mathbf{COV}_{\tilde{\mathbf{v}}}$ can be calculated as:

$$\mathbf{COV}_{\tilde{\mathbf{v}}} = E[\tilde{\mathbf{v}}\tilde{\mathbf{v}}^H] = \sigma^2 (\mathbf{L}_{tot}^H \mathbf{\Sigma}_{tot}^{-1} \mathbf{L}_{tot})^{-1} \quad (16)$$

$$= \sigma^2 \left(\sum_{i=1}^M \mathbf{L}_i^H \mathbf{\Sigma}^{-1} \mathbf{L}_i \right)^{-1} \quad (17)$$

$$= \sigma^2 \left(\sum_{i=1}^M \mathbf{H}_i^* \mathbf{R} \mathbf{H}_i \right)^{-1} \quad (18)$$

where $\mathbf{L}_i = \mathbf{U}\mathbf{H}_i$. In the derivation of $\mathbf{COV}_{\tilde{\mathbf{v}}}$, Eq. (17) is found by some matrix manipulation and Eq. (18) is obtained by using the fact that $\mathbf{U}^H \mathbf{\Sigma}^{-1} \mathbf{U} = \mathbf{R}$. This identity can be simply verified by examining matrices defined in Eqs. (8) and (12). Unfortunately, due to the complex structure of $\left(\sum_{i=1}^M \mathbf{H}_i^* \mathbf{R} \mathbf{H}_i\right)^{-1}$ for $M \geq 1$, finding the exact expression of bit error rate (BER) for $M \geq 1$ is not easy. We derive an upper bound on BER by finding an upper bound on the diagonal elements of $\mathbf{COV}_{\tilde{\mathbf{v}}}$ and show that full diversity is achieved. Because \mathbf{R} is positive definite, for every $1 \leq i \leq M$, $\mathbf{H}_i^* \mathbf{R} \mathbf{H}_i$ is also positive definite. Therefore, we can apply the following lemma.

Lemma 1: For n positive definite matrices \mathbf{A}_i , $1 \leq i \leq n$, we have:

$$\left(\sum_{i=1}^n \mathbf{A}_i\right)^{-1} \leq \sum_{i=1}^n \mathbf{A}_i^{-1} \quad (19)$$

where $\mathbf{B} \leq \mathbf{C}$ means that $\mathbf{C} - \mathbf{B}$ is positive semidefinite.

Proof: This lemma is a straightforward result of the following inequality, which can be found in [12].

$$(\mathbf{A} + \mathbf{B})^{-1} \leq \mathbf{A}^{-1} \quad \mathbf{A}, \mathbf{B} : \text{positive definite matrices}$$

As a result, we can conclude that $\mathbf{COV}_{\tilde{\mathbf{v}}} \leq \sigma^2 \sum_{i=1}^M (\mathbf{H}_i^* \mathbf{R} \mathbf{H}_i)^{-1}$. This inequality implies that the diagonal elements of the covariance matrix of noise are upper bounded as follows:

$$\mathbf{COV}_{\tilde{\mathbf{v}}}(i, i) \leq \frac{\sigma^2 \mathbf{R}^{-1}(i, i)}{\sum_{j=1}^M |h_{(1+(i-1)\bmod K), j}|^2} \quad 1 \leq i \leq NK \quad (20)$$

where $h_{k,m}$ is the channel coefficient between User k and Receive Antenna m .

The BER expression for an AWGN channel with average transmit power of $E[|b_k(i)|^2]$ and noise variance of $\frac{\sigma^2 \mathbf{R}^{-1}(i, i)}{\sum_{j=1}^M |h_{(1+(i-1)\bmod K), j}|^2}$ is equal to:

$$p_i = \frac{\sqrt{\frac{\delta_0 2}{\pi \mathbf{R}^{-1}(i, i)}}}{2 \left(1 + \frac{\delta_0 2}{\mathbf{R}^{-1}(i, i)}\right)^{M+\frac{1}{2}}} \times \frac{\Gamma(M + \frac{1}{2})}{\Gamma(M + 1)} \times {}_2F_1\left(1, M + \frac{1}{2}; M + 1; \frac{1}{1 + \frac{\delta_0 2}{\mathbf{R}^{-1}(i, i)}}\right) \quad (21)$$

where $\delta_0 = \frac{E[|b_k(i)|^2]}{\sigma^2}$. The details of derivation can be found in Appendix A. Due to having the same average transmit power and a lower noise variance, we conclude that BER for each subchannel is upper bounded by p_i , i.e., $P_i \leq p_i$. If we define $D_i = -\lim_{\delta_0 \rightarrow \infty} \frac{\log P_i}{\log \delta_0}$ and $d_i = -\lim_{\delta_0 \rightarrow \infty} \frac{\log p_i}{\log \delta_0}$, it is clear that $D_i \geq d_i$. By using the fact that the hypergeometric function of form ${}_2F_1(1, m + \frac{1}{2}; m + 1; \frac{1}{1+c})$ converges to one as c grows large [13], we can calculate that $d_i = M$. Therefore, the diversity of the i th subchannel is greater than or equal to M . On the other hand, M is the maximum available diversity for this system, which completes the proof of achieving full diversity, i.e. $D_i = M$.

A. Analyzing Asymptotic Performance for Large Number of Receive Antennas

In this section, we show that when M goes to infinity, the correlation between noise samples and the effect of fading coefficients vanishes. For normalizing purposes, instead of using Eq. (15), we perform ZF by multiplying output samples by $\sqrt{M}(\mathbf{L}_{tot}^H \mathbf{\Sigma}_{tot}^{-1} \mathbf{L}_{tot})^{-1} \mathbf{L}_{tot}^H \mathbf{\Sigma}_{tot}^{-1}$ in this section. Then, the inverse of noise covariance will be equal to $\frac{1}{\sigma^2 M} \sum_{i=1}^M \mathbf{H}_i^* \mathbf{R} \mathbf{H}_i$, which can be represented as:

$$\mathbf{COV}_{\tilde{\mathbf{v}}}^{-1} = \frac{1}{\sigma^2} \mathbf{R} \circ (\mathbf{J}_N \otimes \tilde{\mathbf{H}}) \quad (22)$$

where \circ is Hadamard product and \mathbf{J}_N is the $N \times N$ all-ones matrix. $\tilde{\mathbf{H}}$ is also defined as follows:

$$\frac{1}{M} \begin{pmatrix} \sum_{m=1}^M |h_{1,m}|^2 & \sum_{m=1}^M h_{1,m} h_{2,m}^* & \dots & \sum_{m=1}^M h_{1,m} h_{K,m}^* \\ \sum_{m=1}^M h_{2,m} h_{1,m}^* & \sum_{m=1}^M |h_{2,m}|^2 & \dots & \sum_{m=1}^M h_{2,m} h_{K,m}^* \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{m=1}^M h_{K,m} h_{1,m}^* & \sum_{m=1}^M h_{K,m} h_{2,m}^* & \dots & \sum_{m=1}^M |h_{K,m}|^2 \end{pmatrix}$$

Then, by using the law of large numbers, it can be shown that $\tilde{\mathbf{H}} \rightarrow \mathbf{I}_K$ as $M \rightarrow \infty$ [14]. The immediate result is that $\mathbf{COV}_{\tilde{\mathbf{v}}}^{-1}$ and $\mathbf{COV}_{\tilde{\mathbf{v}}}$ approaches to $\frac{1}{\sigma^2} \mathbf{I}_{NK}$ and $\sigma^2 \mathbf{I}_{NK}$, respectively. This result implies that at the output of the ZF receiver, noise samples are independent and SNR for each subchannel is a fixed value of $M\delta_0$, independent of channel coefficients. In other words, the effect of fading coefficients and correlation between noise samples will vanish.

B. Effect of Time Delays on Performance

In this section, we calculate the optimal values of delays for the ZF detection in order to achieve the lowest average BER with one receive antenna at high SNR. Because for $M = 1$ the inequality in Eq. (20) turns into equality, the exact BER expression for each subchannel can be obtained as:

$$P_i = \frac{\sqrt{\frac{\delta_0 2}{\pi \mathbf{R}^{-1}(i, i)}}}{2 \left(1 + \frac{\delta_0 2}{\mathbf{R}^{-1}(i, i)}\right)^{3/2}} \frac{\Gamma(3/2)}{\Gamma(2)} {}_2F_1\left(1, 3/2; 2; \frac{1}{1 + \frac{\delta_0 2}{\mathbf{R}^{-1}(i, i)}}\right)$$

Approximating P_{avg} at high SNR for one receive antenna results in: (see Appendix B for more details)

$$\widetilde{P_{avg}} = \frac{1}{4\sqrt{\pi}NK} \frac{\Gamma(3/2)}{\Gamma(2)} \times \frac{\sum_i \mathbf{R}^{-1}(i, i)}{\delta_0},$$

For a fixed number of users and frame length, in order to maximize $\widetilde{P_{avg}}$, we need to maximize the trace(\mathbf{R}^{-1}) which is related to time delays between different users. In what follows, we derive the relationship between the trace(\mathbf{R}^{-1}) and time delays, and consequently find optimum time delays.

Lemma 2: the sum of the diagonal elements of the inverse of matrix \mathbf{R} is equal to:

$$\text{trace}(\mathbf{R}^{-1}) = \frac{(N-1)(N+1)}{3(1+\tau_1-\tau_K)} + \frac{2N+1}{3(N+1+\tau_1-\tau_K)} + \frac{N(N+2)}{3} \sum_{i=1}^{K-1} \frac{1}{\tau_{i+1}-\tau_i} \quad (23)$$

The proof is presented in Appendix C.

Theorem 1: The optimum time delays which result in the lowest average BER for ZF detection at high SNR are: (τ_1 is assumed to be zero)

$$\tau_{i-1} = \frac{i-2}{i-1} \times \tau_i \quad 3 \leq i \leq K \quad (24)$$

Also τ_K is found by solving the following equation:

$$A\tau_K^4 + B\tau_K^3 + C\tau_K^2 + D\tau_K + E = 0 \quad (25)$$

where

$$A = (1 - (K-1)^2) \frac{(N+2)}{3}.$$

$$B = \frac{-2}{3} (1 - (K-1)^2) N^2 + 2(4(K-1)^2 - 1) \frac{(N+1)}{3}.$$

$$C = \frac{1}{3} (1 - (K-1)^2) N^3 + \frac{2}{3} (1 - 4(K-1)^2) N^2 - 2(K-1)^2 (3N+2).$$

$$D = \frac{2}{3} (K-1)^2 (N^3 + 5N^2 + 8N + 4).$$

$$E = -\frac{1}{3} (K-1)^2 (N^3 + 4N^2 + 5N + 2).$$

The proof is easily obtained by taking the derivation of Eq. (23) with respect to time delays.

For $K = 2$, A will be zero and Eq. (25) is a polynomial of degree 3 which has a closed-form solution as follows:

$$\tau_{opt} = \frac{N+2 - \sqrt[3]{N^3 + 1.5N^2 - 1.5N - 1}}{3} \quad (26)$$

where N is the block length. However, for other values of K , Eq. (25) should be solved numerically. After finding τ_K , the remaining time delays are calculated recursively using Eq. (24). The optimum delay values for different K and N values are reported in Tables I and II. Optimum time delays approach

TABLE I: Optimum Time Delays when $K = 2$

Case	N=10	N=32	N=64	N=128	$N \rightarrow \infty$
K=2	0.5240	0.5077	0.5039	0.5019	0.5

TABLE II: Optimum delays when $N = 128$

Case	N=128
K=4	[0.2505, 0.5010, 0.7514]
K=6	[0.1669, 0.3338, 0.5006, 0.6675, 0.8344]
K=8	[0.1251, 0.2502, 0.3754, 0.5004, 0.6256, 0.7507, 0.8758]

uniform time delays, i.e., $\tau_k = \frac{k-1}{K}$, $2 \leq k \leq K$, as N increases. The effects of time delay values on the performance are studied numerically in the following section.

V. SIMULATION RESULTS

In this section, we provide simulation results in order to validate our theoretical results and compare different methods. In all simulations, channel coefficients are independent Rayleigh fading with variance one, fixed during

the block and changing independently for each block. All users have the same average power of one and variance of noise (σ^2) is equal to $10^{\frac{-\text{SNR}}{10}}$ where SNR is in dB. To avoid inter-block interference, the last symbol of each block should be idle for asynchronous methods. This will reduce spectral efficiency, but it is negligible for large block lengths. In all simulations, the block length is 128 and the time delays are uniform except in the case where we report the time delays to study their effects on the performance. The number of users and the number of receive antennas is denoted by K and M , respectively. When M is not specified, the assumption is that only one receive antenna is used. Transmitted symbols are chosen from BPSK modulation and the comparing criterion is the average bit error rate among all the users.

In Fig. 2, we compare the performance of the asynchronous MLSD method with that of the synchronous ML. Asynchronous MLSD outperforms synchronous ML detection with similar complexity. Fig. 2 also includes the single-user bound for a better comparison. As can be seen in the figure, asynchronous MLSD for $K = 2$ achieves performance of the single user system at high SNR.

Fig. 3 shows the performance of different SIC methods

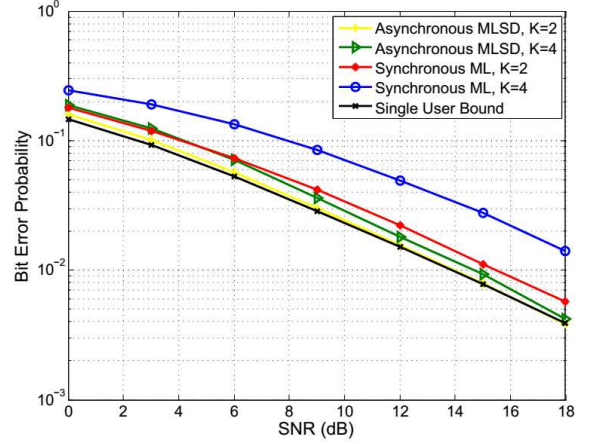


Fig. 2: Comparing asynchronous MLSD and synchronous ML

presented in Section III-B. Our new forward backward belief propagation method using the sampling method in Fig. 1b improves the performance of traditional SIC method by about 3 dB.

Fig. 4 compares the performance of the synchronous and asynchronous ZF detectors. Although asynchronous ZF is possible with one receive antenna, for fair comparison, we consider the cases where the number of receive antennas and users are the same. Since all users are assumed to have the same transmit power, synchronous ZF for $(K = 2, M = 2)$ and $(K = 4, M = 4)$ provides the same performance and both of them have diversity of one. However, for asynchronous ZF detection, diversity of 2 and 4 is achieved for $(K = 2, M = 2)$ and $(K = 4, M = 4)$, respectively. This is due to the sampling diversity as discussed earlier.

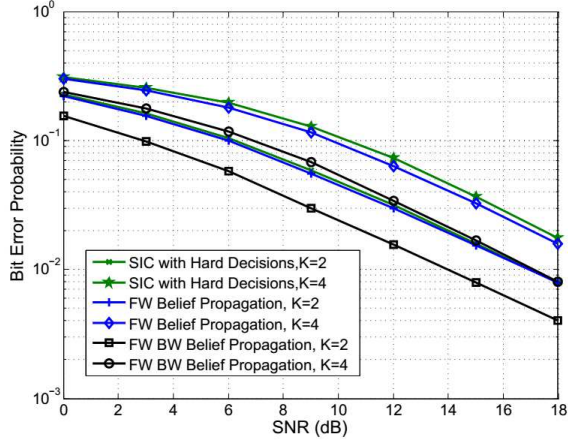


Fig. 3: Performance of SIC method with hard decisions and soft decisions

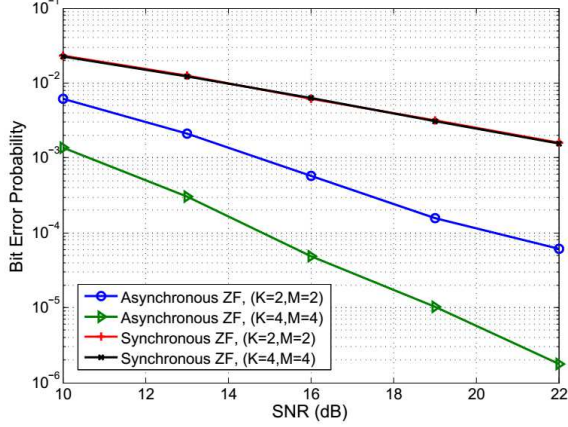


Fig. 4: Comparing synchronous and asynchronous ZF

We study the effects of time delay values on the performance of a ZF system with $K = 4$ users and one receive antenna in Fig. 5. Note that a synchronous ZF solution does not exist in this case as we need at least $M = 4$ receive antennas. We show the results for six different sets of time delays. For optimum time delays we use the result of Section IV-B as reported in Table II. The curve associated with random time delays represents the average performance over uniformly distributed random time delays. The remaining sets of time delays are specified in the figure. The optimum time delays and time delays of $[0.01, 0.1, 0.9]$ have the best and worst performances, respectively. They also have the lowest and the highest $\text{trace}(\mathbf{R}^{-1})$, respectively, which are presented along with other sets of time delays in Table III. As can be seen, a lower $\text{trace}(\mathbf{R}^{-1})$ results in a better performance. This observation is in line with the analysis in Section IV-B where $\text{trace}(\mathbf{R}^{-1})$ was introduced as a criterion to compare the performance of different time delays.

Finally, to compare different methods with each other, we include the performance of all detection methods for $K = 2$

TABLE III: Comparing $\text{trace}(\mathbf{R}^{-1})$ for different time delays in Fig. 5

Time delays	$\text{trace}(\mathbf{R}^{-1})$
$[0.2505, 0.5010, 0.7514]$	8.8404×10^4
$[0.4, 0.6, 0.8]$	9.6639×10^4
$[0.1, 0.4, 0.7]$	1.1065×10^5
$[0.1, 0.2, 0.9]$	1.7347×10^5
$[0.01, 0.1, 0.9]$	6.7784×10^5

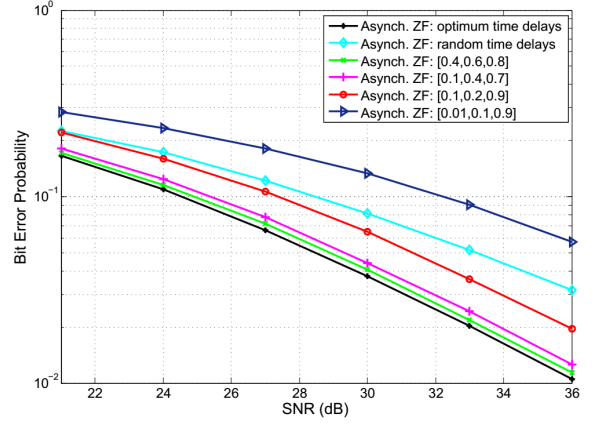


Fig. 5: Effect of time delays in asynchronous ZF detection for $K = 4$

in Fig. 6. Both MLSD and forward-backward BP detection methods not only outperform the synchronous ML detection, but also achieve the performance of the single user system. In addition, the low complexity method of SIC with hard decisions also provides good performance.

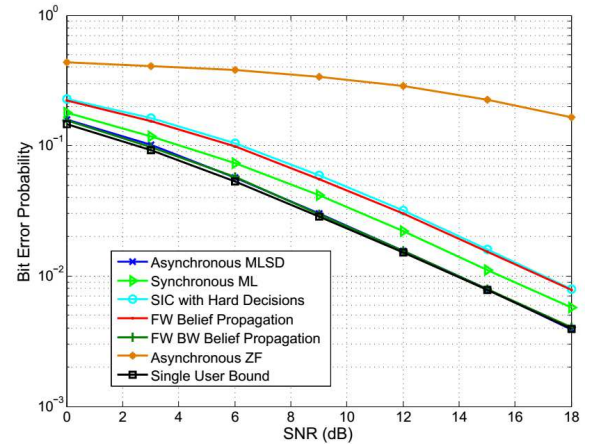


Fig. 6: Comparison of all detection methods for $K = 2$

VI. CONCLUSION

In this paper, we studied benefits of asynchrony when multiple users are sending data simultaneously to a common

receiver. Instead of treating asynchrony as a disruptive factor, we exploited it as an additional resource to cancel interference. We have shown that asynchrony between data streams adds a favorable ISI which makes interference cancellation possible. It also introduces memory to the system which can be exploited by methods like maximum-likelihood sequence detection. In addition to MLSD, a novel forward-backward belief propagation detection method was presented and this method outperforms synchronous ML detection. Exact BER expression for ZF detection was derived and it was verified that a diversity equal to the number of receive antennas is achievable by asynchronous transmission.

REFERENCES

- [1] M. L. Honig, *Advances in multiuser detection*. Wiley Online Library, 2009.
- [2] S. Verdú, *Multiuser detection*. Cambridge University Press, 1998.
- [3] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Communications Magazine*, vol. 34, no. 10, pp. 124–136, 1996.
- [4] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Applications of space-time block codes and interference suppression for high capacity and high data rate wireless systems," *Record of the 32nd Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp. 1803–1810, 1998.
- [5] J. Kazemitarab and H. Jafarkhani, "Multiuser interference cancellation and detection for users with more than two transmit antennas," *IEEE Transactions on Communications*, vol. 56, no. 4, pp. 574–583, 2008.
- [6] S. Shao, Y. Tang, T. Kong, K. Deng, and Y. Shen, "Performance analysis of a modified v-blast system with delay offsets using zero-forcing detection," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6, pp. 3827–3837, 2007.
- [7] A. Das and B. D. Rao, "MIMO Systems with Intentional Timing Offset," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–14, 2011.
- [8] S. Poorkasmaei and H. Jafarkhani, "Asynchronous Orthogonal Differential Decoding for Multiple Access Channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 481–493, 2015.
- [9] M. Avendi and H. Jafarkhani, "Differential distributed space-time coding with imperfect synchronization in frequency-selective channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 1811–1822, 2015.
- [10] S. Verdú, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 35, no. 4, pp. 733–751, 1989.
- [11] J. G. Proakis, *Intersymbol interference in digital communication systems*. Wiley Online Library, 2001.
- [12] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012.
- [13] R. Xu and F. Lau, "Performance analysis for MIMO systems using zero forcing detector over fading channels," *IEEE Proceedings-Communications*, vol. 153, no. 1, pp. 74–80, 2006.
- [14] H. Cramér, *Random variables and probability distributions*. Cambridge University Press, 2004.
- [15] T. Eng and L. B. Milstein, "Coherent DS-CDMA performance in Nakagami multipath fading," *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 1134–1143, 1995.
- [16] M. Dow, "Explicit inverses of Toeplitz and associated matrices," *ANZIAM Journal*, vol. 44, pp. 185–215, 2008.
- [17] D. S. Bernstein, *Matrix mathematics: theory, facts, and formulas*. Princeton University Press, 2009.

APPENDIX A

DERIVATION OF BIT ERROR RATE (BER) EXPRESSION

For an AWGN channel with an average transmit power of $E[|b_k(i)|^2]$ and noise variance of $\frac{\sigma^2 \mathbf{R}^{-1}(i,i)}{\sum_{j=1}^M |h_{(1+(i-1)\text{mod}K),j}|^2}$, the post SNR at the receiver can be expressed as:

$$\delta_i = \frac{\delta_0 \sum_{j=1}^M |h_{(1+(i-1)\text{mod}K),j}|^2}{\mathbf{R}^{-1}(i,i)} \quad (27)$$

where $\delta_0 = \frac{E[|b_k(i)|^2]}{\sigma^2}$. We know that $|h_{i,j}|^2$ follows a chi-squared distribution with two degrees of freedom for all i s and j s. Therefore, $\sum_{j=1}^M |h_{(1+(i-1)\text{mod}K),j}|^2$ is chi-squared distributed with $2M$ degrees of freedom. As a result, the distribution of δ_i can be calculated as follows:

$$P_{\delta_i}(\delta) = \frac{\mathbf{R}^{-1}(i,i)}{\delta_0} \frac{\left(\frac{\mathbf{R}^{-1}(i,i)}{\delta_0} \delta\right)^{M-1} \exp\left(-\frac{\mathbf{R}^{-1}(i,i)}{\delta_0} \delta\right)}{2^M \Gamma(M)} \quad (28)$$

where $\Gamma(\cdot)$ is the Gamma function. For a specific value of SNR, BER varies according to the modulation. We assume that BPSK is used, however, extension to other modulations is straightforward. Based on this assumption, the BER for a given value of SNR, e.g., δ is equal to $Q(\sqrt{2\delta})$. The next step is to calculate the following integral:

$$p_i = \int_0^\infty Q(\sqrt{2\delta}) P_{\delta_i}(\delta) d\delta$$

The integral of $\frac{a^m}{\Gamma(m)} \int_0^\infty \exp(-az) z^{m-1} Q(\sqrt{bz}) dz$ has a closed-form of:

$$\frac{\sqrt{b/2\pi a}}{2 \left(1 + \frac{b}{2a}\right)^{m+1/2}} \frac{\Gamma(m+1/2)}{\Gamma(m+1)} {}_2F_1\left(1, m + \frac{1}{2}; m+1; \frac{1}{1 + \frac{b}{2a}}\right)$$

where ${}_2F_1(q, w; e; r)$ is the hypergeometric function [15]. Therefore, the bit error rate, i.e., p_i is equal to:

$$p_i = \frac{\sqrt{\frac{\delta_0 2}{\pi \mathbf{R}^{-1}(i,i)}}}{2 \left(1 + \frac{\delta_0 2}{\mathbf{R}^{-1}(i,i)}\right)^{M+1/2}} \times \frac{\Gamma(M + \frac{1}{2})}{\Gamma(M+1)} \times {}_2F_1\left(1, M + \frac{1}{2}; M+1; \frac{1}{1 + \frac{\delta_0 2}{\mathbf{R}^{-1}(i,i)}}\right) \quad (29)$$

APPENDIX B

AVERAGE BER AND ITS APPROXIMATION AT HIGH SNR

In Eq. (29), p_i depends on $\mathbf{R}^{-1}(i,i)$ which varies for different values of i , and therefore each subchannel has a different BER. This is unlike the synchronous ZF, where all resulting subchannels have the same performance. In order to evaluate the performance of the entire system, we define the average BER performance as follows:

$$p_{avg} = \frac{\sum_{i=1}^{NK} p_i}{NK} \quad (30)$$

Since p_{avg} is not tractable, we approximate it at high SNR, using the fact that ${}_2F_1(1, m + \frac{1}{2}; m+1; \frac{1}{1+c})$ converges to one as c grows large [13]. Hence, at high SNR, p_{avg} can be approximated as follows:

$$\widetilde{p_{avg}} = \text{Const} \times \frac{\sum_{i=1}^{NK} (\mathbf{R}^{-1}(i,i))^M}{\delta_0^M} \quad (31)$$

where the constant value is equal to $\frac{1}{2^{(M+1)} NK \sqrt{\pi}} \frac{\Gamma(M+\frac{1}{2})}{\Gamma(M+1)}$.

APPENDIX C
PROOF OF LEMMA 2

When the frame length is N , we denote \mathbf{R} by \mathbf{R}^N . Then, we prove by induction that, for all $N \in \mathbb{Z}_+$,

$$\begin{aligned} \text{trace}((\mathbf{R}^N)^{-1}) &= \frac{(N-1)(N+1)}{3(1+\tau_1-\tau_K)} + \frac{2N+1}{3(N+1+\tau_1-\tau_K)} \\ &+ \frac{N(N+2)}{3} \sum_{i=1}^{K-1} \frac{1}{\tau_{i+1}-\tau_i} \end{aligned} \quad (32)$$

Base case: When $N = 1$, \mathbf{R}^1 is equal to \mathbf{R}_{11} which can be written as a generalized Fiedlers matrix whose inverse is given by [16]:

$$\mathbf{R}_{11}^{-1} = -\frac{1}{2} \begin{pmatrix} d_1 & \frac{1}{\tau_2-\tau_1} & \dots & 0 & f \\ \frac{1}{\tau_2-\tau_1} & d_2 & \frac{1}{\tau_3-\tau_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\tau_{K-1}-\tau_{K-2}} & d_{K-1} & \frac{1}{\tau_K-\tau_{K-1}} \\ f & 0 & \dots & \frac{1}{\tau_K-\tau_{K-1}} & d_K \end{pmatrix} \quad (33)$$

where f and d_i s are defined as:

$$f = \frac{1}{\tau_K - \tau_1 - 2} \quad (34)$$

$$d_1 = \frac{1}{\tau_1 - \tau_2} - \frac{1}{\tau_1 - \tau_K + 2} \quad (35)$$

$$d_K = \frac{1}{\tau_7 - \tau_8} - \frac{1}{\tau_1 - \tau_K + 2} \quad (36)$$

$$d_i = \frac{1}{\tau_{i-1} - \tau_i} + \frac{1}{\tau_i - \tau_{i+1}} \quad 2 \leq i \leq K-1 \quad (37)$$

Then, $\text{trace}(\mathbf{R}_{11}^{-1})$ is equal to $(-\frac{1}{2} \sum_{i=1}^K d_i)$, which can be calculated using the above equations:

$$\text{trace}(\mathbf{R}_{11}^{-1}) = \frac{1}{(2 + \tau_1 - \tau_K)} + \sum_{i=1}^{K-1} \frac{1}{\tau_{i+1} - \tau_i} \quad (38)$$

Therefore, Eq. (32) is true for $N = 1$.

Induction step: Suppose Eq. (32) is true for N . We need to show that it also holds for $N + 1$, i.e.,

$$\begin{aligned} \text{trace}((\mathbf{R}^{(N+1)})^{-1}) &= \frac{(N)(N+2)}{3(1+\tau_1-\tau_K)} + \frac{2N+3}{3(N+2+\tau_1-\tau_K)} \\ &+ \frac{(N+1)(N+3)}{3} \sum_{i=1}^{K-1} \frac{1}{\tau_{i+1}-\tau_i} \end{aligned} \quad (39)$$

Because matrix \mathbf{R} follows a recursive structure, \mathbf{R}^{N+1} can be presented as follows:

$$\mathbf{R}^{N+1} = \begin{bmatrix} (\mathbf{R}^N)_{NK \times NK} & (\mathbf{L})_{NK \times K} \\ (\mathbf{L}^T)_{K \times NK} & (\mathbf{R}_{11})_{K \times K} \end{bmatrix}$$

where $\mathbf{L}^T = [\mathbf{0}_{K \times K}, \dots, \mathbf{0}_{K \times K}, (\mathbf{R}_{21})_{K \times K}]$. For calculating the inverse of \mathbf{R}^{N+1} , we use the following lemma for matrix inversion in block form.

Lemma 3: Let na $(m+n) \times (m+n)$ matrix \mathbf{T} be partitioned into a block form:

$$\mathbf{T} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

where the $m \times m$ matrix \mathbf{A} and $n \times n$ matrix \mathbf{D} are invertible. Then, we have:

$$\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{M}^{-1} & -\mathbf{M}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{M}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{M}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}$$

where $\mathbf{M} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ [17].

Here, \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{D} are equal to \mathbf{R}^N , \mathbf{L} , \mathbf{L}^T and \mathbf{R}_{11} , respectively. Therefore, \mathbf{M} is equal to:

$$\mathbf{M} = \mathbf{R}^N - \mathbf{L}(\mathbf{R}_{11})^{-1}\mathbf{L}^T \quad (40)$$

Now, we need to find the inverse of \mathbf{M} . By defining \mathbf{Z} as $(\mathbf{R}^N)^{-1}$, the inverse of \mathbf{M} can be presented as:

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{I}_K & \dots & \mathbf{0}_{K,K} & \mathbf{Z}_{1N}\mathbf{Q}(\mathbf{I}_K - \mathbf{Z}_{NN}\mathbf{Q})^{-1} \\ \mathbf{0}_{K \times K} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \mathbf{I}_K & \mathbf{Z}_{(N-1)N}\mathbf{Q}(\mathbf{I}_K - \mathbf{Z}_{NN}\mathbf{Q})^{-1} \\ \mathbf{0}_{K \times K} & \dots & \mathbf{0}_{K \times K} & (\mathbf{I}_K - \mathbf{Z}_{NN}\mathbf{Q})^{-1} \end{bmatrix} \mathbf{Z} \quad (41)$$

where $\mathbf{Q} = \mathbf{R}_{12}\mathbf{R}_{11}^{-1}\mathbf{R}_{21}$ and \mathbf{Z}_{ij} s are $K \times K$ partitioning blocks of \mathbf{Z} . Also, \mathbf{I}_k and $\mathbf{0}_{i \times j}$ are a $k \times k$ identity matrix and a $i \times j$ all-zero matrix, respectively.

To show the correctness of Eq. (41), we need to take the following steps:

Step 1: By some calculations, it can be shown that $\mathbf{L}(\mathbf{R}_{11})^{-1}\mathbf{L}^T$ is equal to $\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}$. As a result, we have:

$$\mathbf{M} = \mathbf{R}^N - \begin{bmatrix} \mathbf{0}_{(N-1)K \times (N-1)K} & \mathbf{0}_{(N-1)K \times K} \\ \mathbf{0}_{K \times (N-1)K} & \mathbf{Q} \end{bmatrix} \quad (42)$$

Step 2: If we multiply both sides by \mathbf{Z} , we will have:

$$\mathbf{Z}\mathbf{M} = \mathbf{I}_{NK} - \begin{bmatrix} \mathbf{0}_{K \times K} & \dots & \mathbf{0}_{K \times K} & \mathbf{Z}_{1N}\mathbf{Q} \\ \mathbf{0}_{K \times K} & \dots & \mathbf{0}_{K \times K} & \mathbf{Z}_{2N}\mathbf{Q} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{K \times K} & \dots & \mathbf{0}_{K \times K} & \mathbf{Z}_{NN}\mathbf{Q} \end{bmatrix} \quad (43)$$

Step 3: We denote the right hand side of Eq. (43) by \mathbf{X} , then, we can conclude that the inverse of \mathbf{M} is equal to:

$$\mathbf{M}^{-1} = \mathbf{X}^{-1}\mathbf{Z} \quad (44)$$

Step 4: \mathbf{X}^{-1} can be calculated as follows:

$$\mathbf{X}^{-1} = \begin{bmatrix} \mathbf{I}_K & \dots & \mathbf{0}_{K,K} & \mathbf{Z}_{1N}\mathbf{Q}(\mathbf{I}_K - \mathbf{Z}_{NN}\mathbf{Q})^{-1} \\ \mathbf{0}_{K \times K} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \mathbf{I}_K & \mathbf{Z}_{(N-1)N}\mathbf{Q}(\mathbf{I}_K - \mathbf{Z}_{NN}\mathbf{Q})^{-1} \\ \mathbf{0}_{K \times K} & \dots & \mathbf{0}_{K \times K} & (\mathbf{I}_K - \mathbf{Z}_{NN}\mathbf{Q})^{-1} \end{bmatrix} \quad (45)$$

Step 5: Finally, if we plug \mathbf{X}^{-1} in Eq. (44), we will reach Eq. (41).

If we denote $K \times K$ diagonal blocks of \mathbf{M}^{-1} as $[\mathbf{M}^{-1}]_{i,i}$, $1 \leq i \leq N$, then, by use of Lemma 3, $\text{trace}((\mathbf{R}^{N+1})^{-1})$ can be written as:

$$\begin{aligned} \text{trace}((\mathbf{R}^{N+1})^{-1}) &= \sum_{i=1}^N \text{trace}([\mathbf{M}^{-1}]_{i,i}) + \\ &+ \text{trace}(\mathbf{R}_{11}^{-1} + \mathbf{R}_{11}^{-1} \mathbf{R}_{21} [\mathbf{M}^{-1}]_{N,N} \mathbf{R}_{12} \mathbf{R}_{11}^{-1}) \end{aligned} \quad (46)$$

By simplifying Eq. (41), diagonal blocks of \mathbf{M}^{-1} can be presented as follows:

$$1 \leq i \leq N-1 : \quad [\mathbf{M}^{-1}]_{i,i} = \mathbf{Z}_{ii} + \mathbf{Z}_{iN} \mathbf{Q} (\mathbf{I} - \mathbf{Z}_{NN} \mathbf{Q})^{-1} \mathbf{Z}_{Ni} \quad (47)$$

$$i = N : \quad [\mathbf{M}^{-1}]_{i,i} = (\mathbf{I} - \mathbf{Z}_{NN} \mathbf{Q})^{-1} \mathbf{Z}_{NN} \quad (48)$$

In Eq. (46), we set the diagonal blocks of \mathbf{M}^{-1} as Eqs. (47) and (48). Then, by some manipulations, $\text{trace}((\mathbf{R}^{N+1})^{-1})$ can be presented as:

$$\begin{aligned} \text{trace}((\mathbf{R}^{N+1})^{-1}) &= \text{trace}((\mathbf{R}^N)^{-1}) + \text{trace}(\mathbf{R}_{11}^{-1}) \\ &+ \sum_{i=1}^{N-1} \text{trace}(\mathbf{Z}_{iN} \mathbf{Q} (\mathbf{I} - \mathbf{Z}_{NN} \mathbf{Q})^{-1} \mathbf{Z}_{Ni}) \\ &+ \text{trace}((\mathbf{I} - \mathbf{Z}_{NN} \mathbf{Q})^{-1} \mathbf{Z}_{NN}) - \text{trace}(\mathbf{Z}_{NN}) \\ &+ \text{trace}(\mathbf{R}_{11}^{-1} \mathbf{R}_{21} (\mathbf{I} - \mathbf{Z}_{NN} \mathbf{Q})^{-1} \mathbf{Z}_{NN} \mathbf{R}_{12} \mathbf{R}_{11}^{-1}) \end{aligned} \quad (49)$$

The first and second terms in Eq. (49) can be calculated by induction hypothesis and induction base, respectively. Calculating other terms in Eq. (49) is tedious but similar for different values of K . Therefore, we only calculate it for $K = 2$ and skip the rest. For $K = 2$, \mathbf{Q} is equal to:

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1-\tau}{1+\tau} \end{bmatrix} \quad (50)$$

where $\tau = \tau_2 - \tau_1$.

If we plug $\mathbf{Q} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1-\tau}{1+\tau} \end{bmatrix}$ in Eq. (49), after some calculations we will have:

$$\begin{aligned} \text{trace}((\mathbf{R}^{N+1})^{-1}) &= \text{trace}((\mathbf{R}^N)^{-1}) + \frac{2}{1 - (1-\tau)^2} \\ &+ \frac{1-\tau}{(1+\tau) - (1-\tau)r(2N, 2N)} \sum_{i=1}^{2N} (r(2N, i))^2 \\ &+ \frac{(1+\tau)(1+(\tau-1)^2)}{(2-\tau)^2[(1+\tau) - (1-\tau)r(2N, 2N)]} r(2N, 2N) \end{aligned} \quad (51)$$

where $r(i, j)$ is the (i, j) th element of matrix $(\mathbf{R}^N)^{-1}$. By induction hypothesis, the first term in Eq. (51) is equal to $\frac{(N-1)(N+1)}{3(1-\tau_2+\tau_1)} + \frac{N(N+2)}{3(\tau_2-\tau_1)} + \frac{2N+1}{3(N+1-\tau_2+\tau_1)}$. For calculating Eq. (51), we also need values of $r(2N, i)$, $1 \leq i \leq 2N$, which are elements of the last row of $(\mathbf{R}^N)^{-1}$. Due to the special

structure of matrix \mathbf{R} , values of $r(2N, i)$ can be calculated as follows:

$$\begin{cases} r(2N, 2i-1) = \frac{\tau-i}{\tau(N+1-\tau)} & 1 \leq i \leq N \\ r(2N, 2i) = \frac{i}{\tau(N+1-\tau)} \end{cases} \quad (52)$$

To verify Eq. (52), we can multiply the last row of $(\mathbf{R}^N)^{-1}$, i.e., $[r(2N, 1), r(2N, 2), \dots, r(2N, 2N)]$, by different columns of \mathbf{R}^N as follows:

$$\text{1st column: } \frac{(\tau-1)1}{\tau(N+1-\tau)} + \frac{(1)(1-\tau)}{\tau(N+1-\tau)} = 0$$

$$(2i)_{th} \text{ column: } 1 \leq i \leq N-1$$

$$\frac{(\tau-i)(1-\tau)}{\tau(N+1-\tau)} + \frac{(i)1}{\tau(N+1-\tau)} + \frac{(\tau-(i+1))\tau}{\tau(N+1-\tau)} = 0$$

$$(2i-1)_{th} \text{ column: } 2 \leq i \leq N$$

$$\frac{((i-1))(\tau)}{\tau(N+1-\tau)} + \frac{(\tau-i)1}{\tau(N+1-\tau)} + \frac{(i)(1-\tau)}{\tau(N+1-\tau)} = 0$$

$$2N_{th} \text{ column: } \frac{(\tau-N)(1-\tau)}{\tau(N+1-\tau)} + \frac{(N)1}{\tau(N+1-\tau)} = 1$$

These results verify that the last row of $(\mathbf{R}^N)^{-1}$ follows the pattern in Eq. (52).

The last step is to plug Eq. (52) into Eq. (51). As a result, $\text{trace}((\mathbf{R}^{N+1})^{-1})$ is equal to $\frac{(N)(N+2)}{3(1-\tau_2+\tau_1)} + \frac{(N+1)(N+3)}{3(\tau_2-\tau_1)} + \frac{2N+3}{3(N+2-\tau_2+\tau_1)}$, which verifies the induction step and completes the proof.

This figure "BP.png" is available in "png" format from:

<http://arxiv.org/ps/1609.07032v1>

This figure "MLSD.png" is available in "png" format from:

<http://arxiv.org/ps/1609.07032v1>

This figure "ZF.png" is available in "png" format from:

<http://arxiv.org/ps/1609.07032v1>

This figure "all.png" is available in "png" format from:

<http://arxiv.org/ps/1609.07032v1>

This figure "asynch.png" is available in "png" format from:

<http://arxiv.org/ps/1609.07032v1>

This figure "asynch2.png" is available in "png" format from:

<http://arxiv.org/ps/1609.07032v1>

This figure "time_delay.png" is available in "png" format from:

<http://arxiv.org/ps/1609.07032v1>